AD-A061 838   SACLANT ASW RESEARCH CENTRE LA SPEZIA (ITALY)   F/G 12/1

A COMPARISON OF METHODS FOR OPTIMUM MATCHING OF THE MEAN VALUES--ETC(U)

JUN 78 A MACK, M J DAINTITH

UNCLASSIFIED   SACLANTCEN-SM-109   NL

1 OF 1

AD
A061838

END

DATE
FILMED

2-79

DDC

LEVEL II

12

14

SACLANTCEN Memorandum SM - 109

SACLANTCEN Memorandum SM - 109

SACLANT ASW RESEARCH CENTRE MEMORANDUM

AD A061838

6

A COMPARISON OF METHODS FOR OPTIMUM MATCHING OF THE MEAN VALUES OF TWO SAMPLES

by

10

ANGELA MACK AND MAURICE J. DAINTITH

DDC RECEIVED NOV 17 1978 F

11

15 JUNE 1978

12 40p.

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

NORTH ATLANTIC TREATY ORGANIZATION

LA SPEZIA, ITALY

DDC FILE COPY

78 11 16 200

312 950

LB

Published by

STI

SACLANTCEN

INITIAL DISTRIBUTION

MINISTRIES OF DEFENCE

	Copies
MOD Belgium	2
DND Canada	10
CHOD Denmark	8
MOD France	8
MOD Germany	15
MOD Greece	11
MOD Italy	10
MOD Netherlands	12
CHOD Norway	10
MOD Portugal	5
MOD Turkey	5
MOD U.K.	16
SECDEF U.S.	61

NATO AUTHORITIES

Defence Planning Committee	3
NAMILCOM	2
SACLANT	10
SACLANTREPEUR	1
CINCWESTLANT/COMOCEANLANT	1
COMIBERLANT	1
CINCEASTLANT	1
COMSUBACLANT	1
COMMAIREASTLANT	1
SACEUR	2
CINCNORTH	1
CINCSOUTH	1
COMNAVSOUTH	1
COMSTRIKFORSOUTH	1
COMEDCENT	1
COMMARAIRMED	1
CINCHAN	1

SCNR FOR SACLANTCEN

	Copies
SCNR Belgium	1
SCNR Canada	1
SCNR Denmark	1
SCNR Germany	1
SCNR Greece	1
SCNR Italy	1
SCNR Netherlands	1
SCNR Norway	1
SCNR Portugal	1
SCNR Turkey	1
SCNR U.K.	1
SCNR U.S.	2
SECGEN Rep.	1
NAMILCOM Rep.	1

NATIONAL LIAISON OFFICERS

NLO Canada	1
NLO Denmark	1
NLO Germany	1
NLO Italy	1
NLO U.K.	1
NLO U.S.	1

NLR TO SACLANT

NLR Belgium	1
NLR Canada	1
NLR Germany	1
NLR Greece	1
NLR Italy	1
NLR Norway	1
NLR Portugal	1
NLR Turkey	1

Total initial distribution	232
SACLANTCEN Library	10
Stock	38
Total number of copies	280

SACLANTCEN MEMORANDUM SM-109

NORTH ATLANTIC TREATY ORGANIZATION

SACLANT ASW Research Centre
Viale San Bartolomeo 400, I-19026 San Bartolomeo (SP), Italy
Tel: (0187)503540
Telex: 28148

A COMPARISON OF METHODS FOR OPTIMUM MATCHING OF THE MEAN VALUES OF TWO SAMPLES

by

Angela Mack and Maurice J. Daintith

15 June 1978

This memorandum has been prepared within the SACLANTCEN Systems Research Division.

L. Whicker

L. WHICKER
Division Chief

TABLE OF CONTENTS

TABLE OF CONTENTS (cont'd)

List of Figures

A COMPARISON OF METHODS FOR OPTIMUM MATCHING OF THE MEAN VALUES OF TWO SAMPLES

by

Angela Mack and Maurice J. Daintith

ABSTRACT

Two sample ranking methods are used to match the means of the two samples taken from a statistical population and results are compared with the random method.

INTRODUCTION

There are many situations in which the predominant need is to match two outputs as accurately as possible, so that by subtraction a zero output is achieved. As an example, consider a split-beam transducer, in which, when the two halves are opposite phase an important feature is the depth of the null. In this situation a small mismatch produces a disproportionate effect upon the null. If we refer levels to that produced by adding the two outputs it is easily shown that (as compared with the theoretical $-\infty$ dB for exact matching), a 1 dB mismatch gives a -26 dB null, a 2 dB mismatch a -19 dB null, a 3 dB mismatch a -15 dB null.

Matching is thus critical, but this may be a time-consuming and expensive process. This paper compares various methods for attaining the best matching that is, at the same time, more economical in time and cost than high quality control or selection, by taking advantage of the fact that we are usually dealing with arrays of elements numbering from the order of tens to (exceptionally) hundreds.

The aim, therefore, is to match the means of two samples taken from a statistical population or in other words 'how can two teams A and B be chosen so that they are as well matched as possible?' There are various methods of doing this:

1) Random: The samples are picked at random from the given distribution.

2) Selection: One item is taken at random from the population. Items are then successively picked until one is found that has the same weight as the first item chosen and subsequently the first item is put in A and the second in B. This is repeated until the samples are of the desired size. This method gives amples with the same total weight.

3) Matching Sums: The samples A and B are picked so that the total weight of the items in A and items in B are equal.

4) Best Matching: If the samples A and B are to be of N/2 items each, a sample of N items is chosen at random. These are then divided into A and B so that the weights (individual or total) are as well matched as possible.

5) Ranking: A sample of N items is taken at random. These are then ordered and the samples A and B are chosen from these taking into account this ordering.

6) Linear Programming: The problem may be turned into a linear program whereby the function to minimize is the difference of the total weights $\Sigma(A_i - B_i)$ and the constraints will depend on the particular problem.

Methods 2 and 3 obviously give the best match but are wasteful, since many items have to be picked and then discarded before completing the samples. They also require many measurements, since each item has to be given a weight and then these weights have to be compared. Method 4 does not give such good results as 2 and 3 but is not wasteful and involves less measurements. Method 5 requires less measurements than 2, 3, and 4 and although results are not so good can produce a definite improvement on 1. In certain cases one may not be able to assign to each item a weight, although using some criterion, they can be ordered. In situations such as these, methods 2, 3, 4, and 6 cannot be used, leaving only random and ranking methods. When sufficient data is available method 6 may seem a good approach, but in fact this may not be so. (See Appendix E.)

This study compares two ranking methods with the random method for two particular parent populations, Rectangular and Gaussian. The two ranking methods considered are:

a. Alternate Ranking

Having taken a sample of N items from the parent population these are ordered

$$x_1 \leq x_2 \leq \ldots\ldots \leq x_N$$

(N is even and indicates the total number of items chosen such that each sample A and B contains N/2 items.)
Starting from x_1, divide into the two samples according to the rule

A B A B A B ....

If Δ represents the difference of the sum of the weights of the items in A and the sum of the weights of those in B then, for any rule based on ranking,

$$\Delta = \sum_{r=1}^{N} \gamma_r x_r , \quad \text{where } \gamma_r = \pm 1.$$

Here

$\gamma_r = 1$ r odd (sample A)

$\gamma_r = -1$ r even (sample B).

b. Bialternate Ranking

Repeat as for the alternate method but divide into the two samples according to the rule

A B B A A B B A

For

$$\Delta = \sum_{1}^{N} \gamma_r \, x_r \, ,$$

where

$$\gamma_r = 1 \qquad \text{if} \quad r \equiv 0 \text{ or } 1 \quad \text{mod } 4$$

$$\gamma_r = -1 \qquad \text{if} \quad r \equiv 2 \text{ or } 3 \quad \text{mod } 4$$

Method b is expected to give better results than a. In fact, in the alternate method, for every pair AB the larger one goes in B and hence the sum of the B's is always greater than the sum of the A's. To compensate for this, in every other pair the order AB is inverted to give AB BA AB BA ... With this method the pair AB puts the larger number in B while the pair BA put the larger one in A , thus reducing the difference between the sum of the A's and the sum of the B's.

1 METHOD

The problem was approached both analytically and numerically. Because the analytic method proved difficult the numerical approach was used both to check results and to obtain them where analytically it was impossible.

1.1 Analytic Solution

For each parent population considered, the requirement is to estimate the mean value $\bar{\Delta}$ and the variance about the origin $\overline{\Delta^2}$ where Δ has already been defined in the Introduction. Considering a generic population of probability distribution $p(x)$ and denoting the cumulative integral $\int_{-\infty}^{x_t} p(x)\,dx$ by $P(x_t)$ and $Q_t = 1 - P_t = \int_{x_t}^{\infty} p(x)\,dx$

the following general expressions are obtained (see Appendices A and B)

$$\bar{\Delta} = N! \sum_{r=1}^{N} \int_{-\infty}^{+\infty} \gamma_r \, y \, p \, \frac{P^{r-1} \, Q^{N-r}}{(r-1)! \, (N-r)!} \, dy \qquad \text{(Eq. 1)}$$

$$(P = P(y); \quad p = p(y))$$

$$\bar{\Delta}^2 = Nv^2 + 2N! \sum_{r=1}^{N-1} \sum_{S=r+1}^{N} \gamma_r \gamma_s \int_{-\alpha}^{+\alpha} y p_y \, dy \int_{-\alpha}^{y} z \, p_z \, dz \, \frac{P_z^{r-1} \, Q_y^{N-s} \, (P_y - P_z)^{s-r-1}}{(r-1)! \, (N-r)! \, (s-r-1)!}$$

(Eq. 2)

$(P_y = p(y) \; ; \; P_z = p(z))$

$v^2 =$ Variance about the origin of the parent population.

1.1.1 Rectangular Distribution

Consider the rectangular distribution

$$p(x) = \begin{cases} 1 & x \in [0, 1] \\ 0 & \text{elsewhere} \end{cases}$$

The population has a mean of a ½ and the variance about the origin is

$$v^2 = \int_0^1 x^2 \, p(x) \, dx = \frac{1}{3} \; .$$

Also

$$P(x) = \int_0^x p(x) \, dx = x \qquad \text{for} \quad x \in [0, 1]$$

$$Q(x) = 1 - x \quad .$$

The limits $\pm\alpha$ are replaced by 1 and 0. Equations (1) and (2) become

$$\bar{\Delta} = N! \sum_{r=1}^{N} \int_0^1 \gamma_r \, y \, \frac{y^{r-1}}{(r-1)!} \frac{(1-y)^{N-r}}{(N-r)!} \, dy$$

$$\bar{\Delta}^2 = \frac{N}{3} + 2N! \sum_{r=1}^{N-1} \sum_{S=r+1}^{N} \gamma_r \gamma_s \int_0^1 y \, dy \int_0^y z \, dz \, \frac{z^{r-1} (1-y)^{N-s} (y-z)^{s-r-1}}{(r-1)! \, (N-r)! \, (s-r-1)!}$$

a. <u>Alternate choice</u>

$$\gamma_r = (-1)^{r-1}$$

$$\bar{\Delta}_A = N! \sum_{r=1}^{N} \int_0^1 y \, \frac{(-y)^{r-1}}{(r-1)!} \, \frac{(1-y)^{N-r}}{(N-r)!} \, dy \quad .$$

The summation is now a binomial, so

$$\bar{\Delta}_A = N \int_0^1 y(1-2y)^{N-1}\, dy \quad .$$

Integrating by parts and taking account of the fact that N is even,

$$\bar{\Delta}_A = \frac{-N}{2(N+1)} \quad .$$

For the variance

$$\overline{\Delta^2_A} = \frac{N}{3} + 2N! \sum_{r=1}^{N-1} \sum_{S=r+1}^{N} (-1)^{r+s} \int_0^1 y\, dy \int_0^y z\, dz\; z^{r-1} \frac{(1-y)^{N-s}(y-z)^{s-r-1}}{(r-1)!(N-r)!(s-r-1)!} \, ,$$

but

$$(-1)^{r+s} = (-1)^{r-s}(-1)^{2s} = (-1)^{r-s} = -(-1)^{r-s-1}$$

$$\Rightarrow \overline{\Delta^2_A} = \frac{N}{3} - 2N! \int_0^1 y\, dy \int_0^y z\, dz \sum_{r=1}^{N-1} \sum_{S=r+1}^{N} z^{r-1} \frac{(1-y)^{N-s}(z-y)^{s-r-1}}{(r-1)!(N-r)!(s-r-1)!}$$

The double summation is a multinomial equal to

$$\frac{1}{(N-2)!} (z + (1-y) + (z-y))^{N-2}$$

$$\Rightarrow \overline{\Delta^2_A} = \frac{N}{3} - 2N(N-1) \int_0^1 y\, dy \int_0^y z\, dz\; (2z-2y+1)^{N-2} \quad .$$

Integrating by parts for both integrals,

$$\Delta^2_A = \frac{N}{4(N+1)} \quad .$$

Using these expressions for $\bar{\Delta}$ and $\overline{\Delta^2}$ the variance about the mean $\bar{\Delta}$ is

$$\sigma^2_A = \overline{\Delta^2_A} - \bar{\Delta}^2_A = \frac{N}{4(N+1)} - \frac{N^2}{4(N+1)^2} = \frac{N}{4(N+1)^2} \quad .$$

2.3 *The Frequency Distributions of $\Delta/N/2$*

Figures 3 to 8 give an idea of the frequency distributions of $\Delta/ N/2$ i.e. the frequency distribution of the difference of the sample means for various values of N ($\bar{\Delta}$ in the alternate ranking method is taken

b. Bialternate Choice

The following results were obtained (Appendix C)

$$N/2 \quad \text{even} \Rightarrow \bar{\Delta}_B = 0 \qquad \bar{\Delta}^2_B = \tfrac{1}{2}(N+1)$$

$$N/2 \quad \text{odd} \Rightarrow \bar{\Delta}_B = \frac{-1}{N+1} \qquad \bar{\Delta}^2_B = \frac{3}{2}(N+1)$$

For the variance about the means

$$N/2 \quad \text{even} \quad \sigma^2_B = \tfrac{1}{2}(N+1)$$

$$N/2 \quad \text{odd} \quad \sigma^2_B = (3N+1)/2(N+1)^2$$

c. Random Choice

Two samples of N/2 items are selected at random since the value of σ^2 for single samples is 1/12 and the variance of the sample mean is $\frac{\sigma^2}{N/2} = \frac{1}{6N}$.

Then the variance of the difference of the mean is

$$2 \times \frac{1}{6N} = \frac{1}{3N}$$

But variance of difference of means $= \bar{\Delta}^2_R / (N/2)^2$

$$\Rightarrow \bar{\Delta}^2_R = \frac{1}{3N} \times \frac{N^2}{4} = \frac{N}{12}$$

The results are summarized in Table 1 and shown graphically in Figs. 1 & 2.

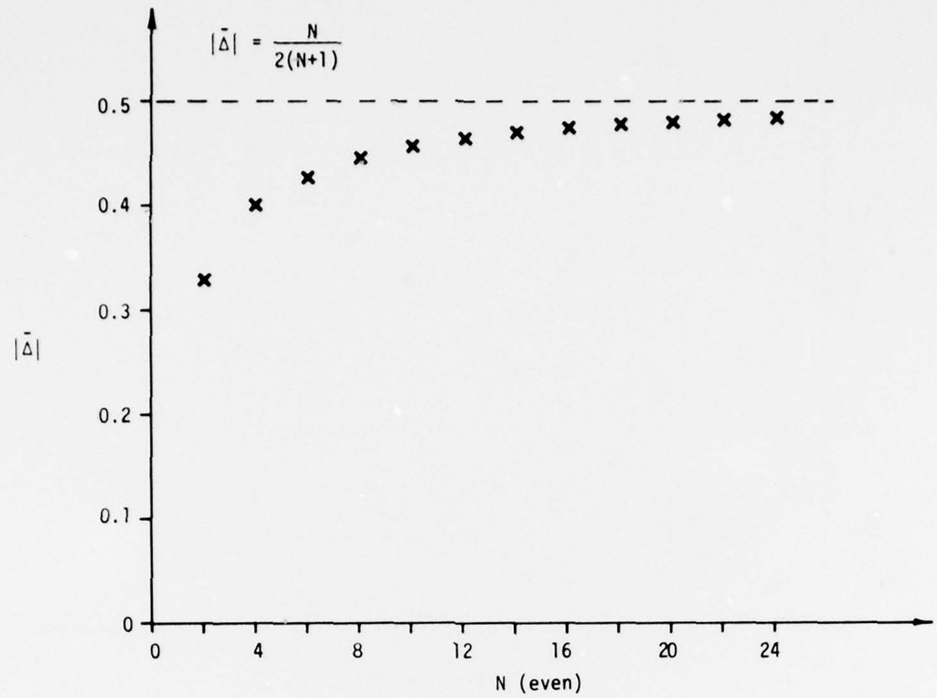
TABLE 1

RECTANGULAR DISTRIBUTION, MEAN ½, VARIANCE ABOUT MEAN 1/12

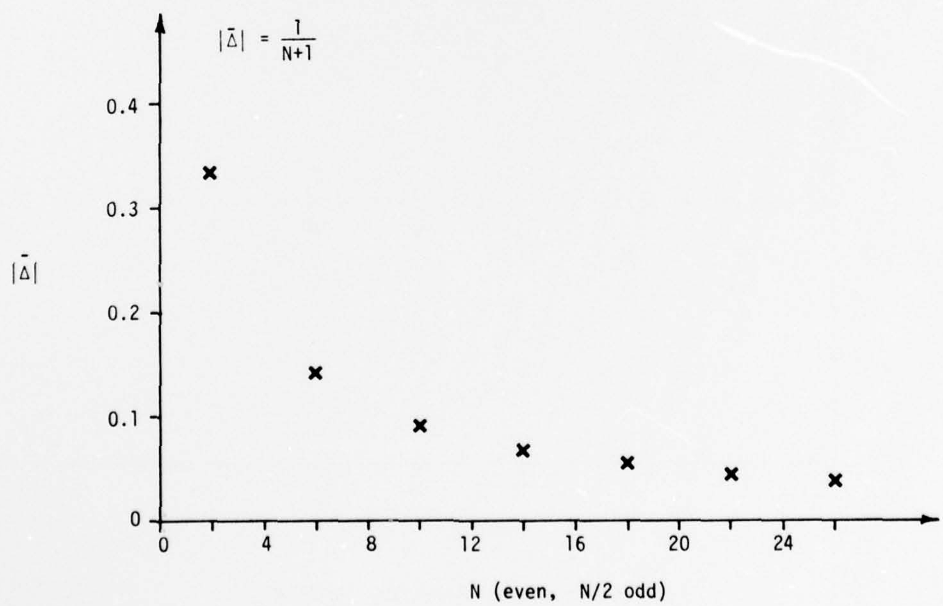
Method	$\bar{\Delta}$	$\bar{\Delta}^2$	$\sigma_{\bar{\Delta}}$	
Random	0	N/12	$\frac{1}{2}\sqrt{\frac{N}{3}}$	N even
Alternate	$-\frac{N}{2(N+1)}$	$\frac{N}{4(N+1)}$	$\frac{1}{N+1}\sqrt{\frac{N}{4}}$	N even
Bialternate	0	$\frac{1}{2(N+1)}$	$\sqrt{\frac{1}{2(N+1)}}$	N even, $\frac{N}{2}$ even
Bialternate	$-\frac{1}{N+1}$	$\frac{3}{2(N+1)}$	$\frac{1}{N+1}\sqrt{\frac{3N+1}{2}}$	N even, $\frac{N}{2}$ odd

An indication of the real value of the standard deviation may be had by calculating confidence intervals. This can be done using χ^2. For confidence intervals with a level of confidence of 95%

$$\frac{S\sqrt{n}}{\quad} < \sigma < \frac{S\sqrt{n}}{\quad}$$

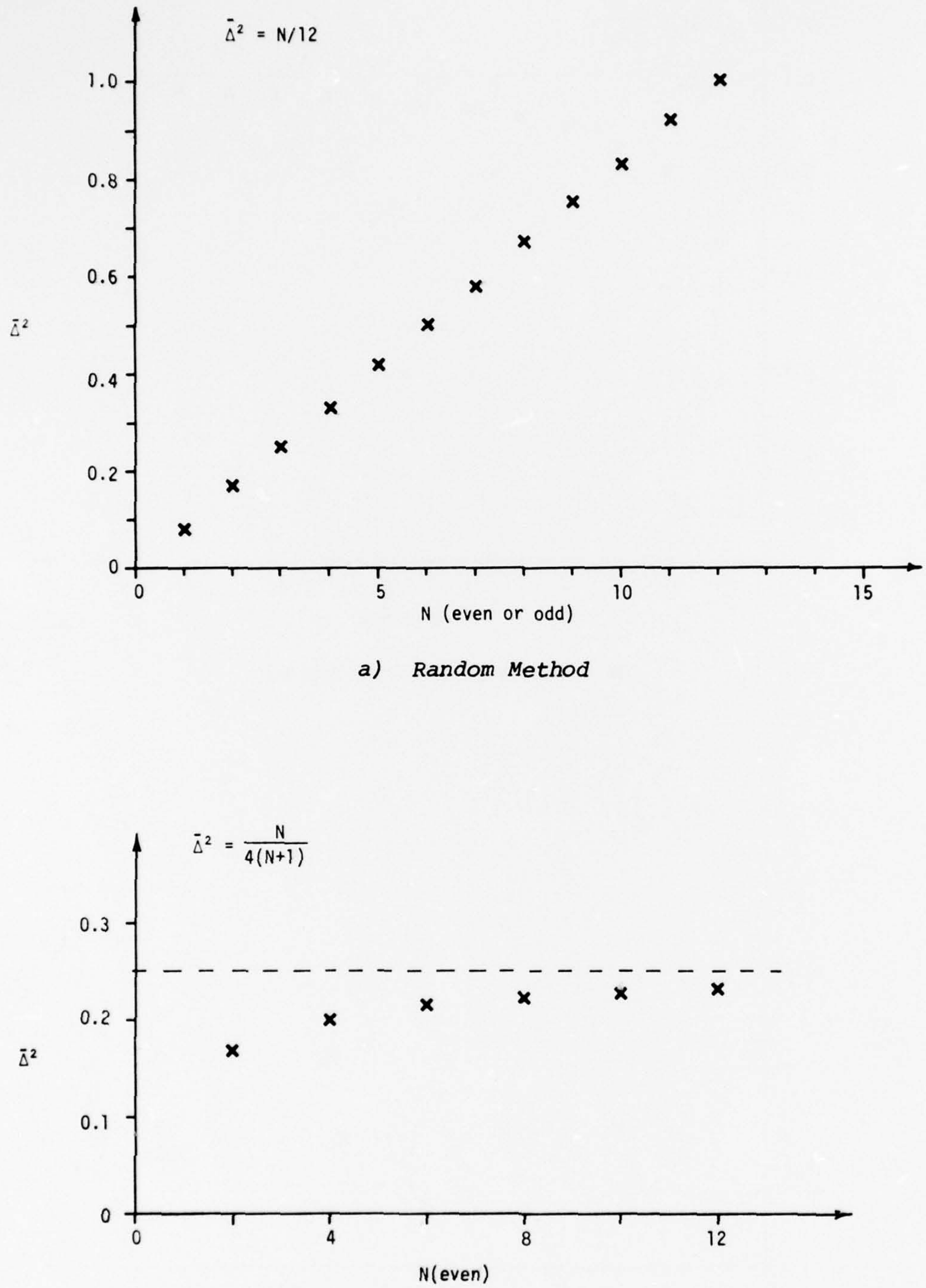


a) Alternate Ranking Method



b) Bialternate Ranking Method

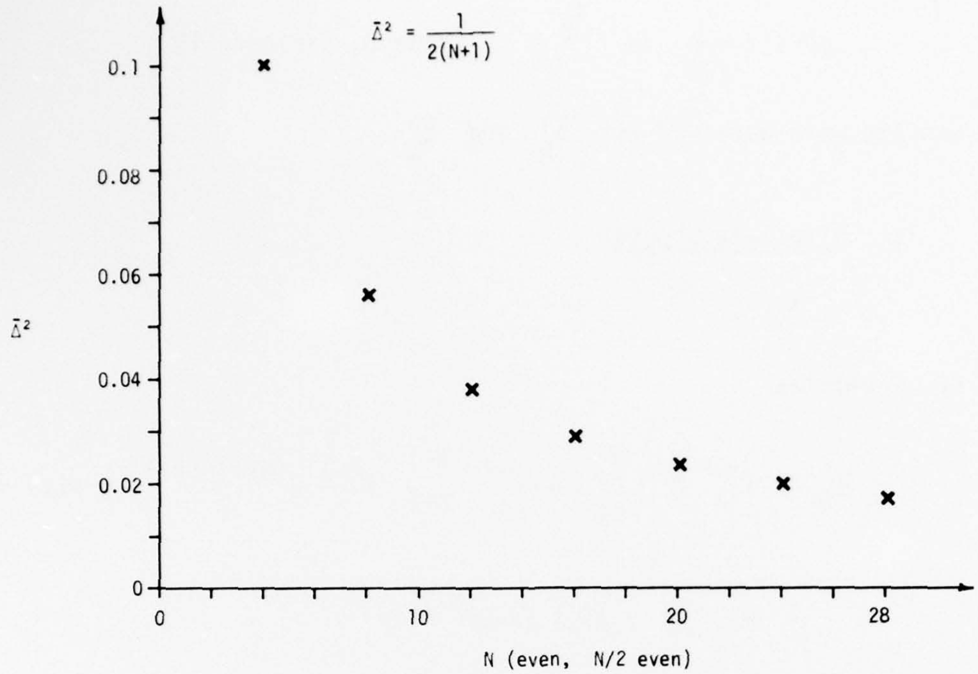
FIG. 1 VARIATION OF $\bar{\Delta}$ WITH N WHERE SAMPLES ARE TAKEN FROM A RECTANGULAR DISTRIBUTION



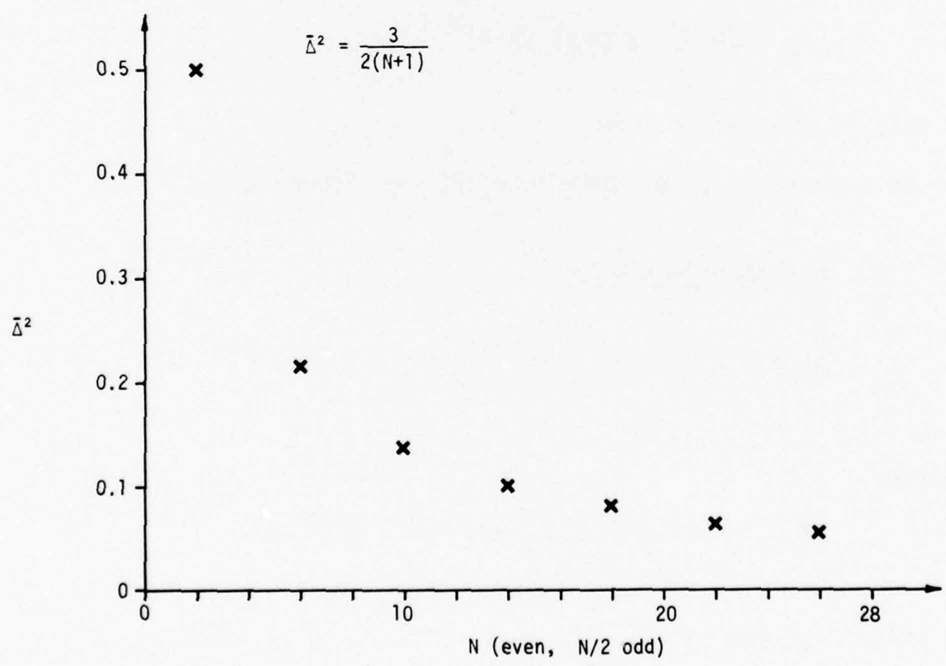
a) Random Method

b) Alternate Ranking Method

FIG. 2 VARIATION OF $\bar{\Delta}^2$ WITH N



c) Bialternate Ranking Method



d) Bialternate Ranking Method

FIG. 2 VARIATION OF $\bar{\Delta}^2$ WITH N

1.1.2 Gaussian Distribution

The distribution is

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \qquad \text{(mean 0, variance 1).}$$

No results were obtained for $\overline{\Delta}^2_A$ and $\overline{\Delta}^2_B$.

a. <u>Alternate Choice</u>

$$\gamma_r = (-1)^{r-1}$$

so Eq. 1 becomes

$$\overline{\Delta}_A = \int_{-\alpha}^{+\alpha} N \sum_{r=1}^{N} (-1)^{r-1} \binom{N-1}{r-1} P^{r-1} Q^{(N-1)-(r-1)} \; y\, p(y)\, dy$$

$$= N \int_{-\alpha}^{+\alpha} y\, p(y)\, (P-Q)^{N-1}\, dy .$$

Now x is an odd function; the distribution, being normal is an even function; $P(-x) = Q(x)$ so $P-Q$ is odd. Hence for N even, the integral is an even function and so

$$\overline{\Delta}_A = 2N \int_{o}^{\alpha} y\, p(y)\, (P-Q)^{N-1}\, dy$$

and this is unequal to zero.

(For an approximation of this integral see Appendix D.)

b. <u>Bialternate Choice</u>

$$\overline{\Delta}_B = N \int_{-\alpha}^{+\alpha} \sum \gamma_r \frac{(N-1)!}{(r-1)!\,(N-r)!} \; y\, p(y)\, P^{r-1} Q^{N-r} \; dy .$$

Consider

$$\int_{-\alpha}^{0} \gamma_r \frac{(N-1)!}{(r-1)!\,(N-r)!} \; y\, p(y)\, P^{r-1} Q^{N-r}\, dy$$

and write $-y$ for y

$$\int_{+\alpha}^{0} \gamma_r \frac{(N-1)!}{(r-1)!(N-r)!} \; y\, p(y)\, Q^{r-1} P^{N-r}\, dy$$

put $N-r = r' - 1$

$$\int_{+\alpha}^{0} \gamma_{N-r'+1} \frac{(N-1)!}{(r'-1)!(N-r')!} y\, p(y)\, P^{r'-1} Q^{N-r'} dy$$

If $N \equiv 0 \mod 4 \Rightarrow \gamma_{N-r'+1} = \gamma_r$

Hence the two integrals are of equal magnitude but opposite sign and so $\bar{\Delta}_B = 0$.

c. Random Choice

The expected value of Δ will obviously be 0. For a normal distribution and a variance of σ^2 the standard deviation of the averages of samples of n items is

$$\sigma_n = \frac{\sigma}{\sqrt{n}}$$

so

$$\frac{\sqrt{\bar{\Delta}_R^2}}{N/2} = \sigma \sqrt{\frac{4}{N}}$$

Here $\sigma = 1$ standard deviation of the difference of the sample means $= \sqrt{\frac{4}{N}}$.

1.2 Numerical Solution

The solution was based on a Monte Carlo method. For the rectangular distribution a random-number generator was used to generate uniformly-distributed random numbers u_i in the range $u_i \in [0, 1]$ using the formula

$$u_i = \text{fractional part of } [(\pi + u_{i-1})^5$$

For the gaussian distribution this was modified to generate pairs of normal random deviates with mean 0 and variance 1 using method 3 on p. 953 of Ref. 1. In each case, between 20 and 30 groups of N random numbers were generated. For the random selection each group was split into two, giving the two samples A and B with N/2 numbers in each. The sample means of A and B were calculated and then the difference of these means. Finally the average and standard deviation of these

differences was found. For the ranking methods the groups of numbers were ordered and then divided into the samples A and B, first using the alternate ranking method and secondly using the bialternate ranking method. As for the random selection, the differences of the sample means were calculated and then the average and standard deviations of these. Tables 2 & 3 indicate the results obtained.

TABLE 2

RESULTS OBTAINED FOR A RECTANGULAR DISTRIBUTION

MEAN ½, VARIANCE 1/12

METHOD	N	MEAN	STANDARD DEVIATION
Random	4	-0.02	0.25
Alternate		-0.22	0.10
Bialternate		0.01	0.14
Random	6	-0.01	0.29
Alternate		-0.16	0.06
Bialternate		-0.04	0.07
Random	20	-2.25×10^{-3}	0.10
Alternate		-0.04	0.02
Bialternate		2.45×10^{-3}	0.02

TABLE 3

RESULTS OBTAINED FOR A GAUSSIAN DISTRIBUTION

MEAN 0, VARIANCE 1

METHOD	N	MEAN	STANDARD DEVIATION
Random	4	0.06	0.61
Alternate		-0.50	0.35
Bialternate		0.01	0.33
Random	8	-0.11	0.49
Alternate		-0.32	0.12
Bialternate		0.01	0.14
Random	20	0.03	0.33
Alternate		-0.14	0.05
Bialternate		-0.01	0.05

Figures 3 to 8 were drawn from the numerical data obtained. For each one a frequency table of the difference of the sample means for a given N was constructed and this was then used to construct the histogram.

2 DISCUSSION

2.1 Comparison of Analytical and Numerical Results

Comparing the numerical and analytical data it must be remembered what has been calculated in each case. Analytically, calling Δ the difference of the sum of the weights of items in A and weights of items in B = $X_1 - X_2$, expressions for

$$\bar{\Delta} = \int_{-\alpha}^{+\alpha} \Delta \; p(\Delta) \; d\Delta \qquad \text{(mean)}$$

and

$$\bar{\Delta}^2 = \int_{-\alpha}^{+\alpha} \Delta^2 \; p(\Delta) \; d\Delta \qquad \text{(variance about origin)}$$

have been found.

Numerically, taking the two samples A and B, $(X_1 - X_2)/N/2$ was calculated and this was repeated for the 20 to 30 sets of samples. Finally the mean and standard deviation of these was calculated:

$$\text{mean} \quad \sim \bar{\Delta} \,/\, \frac{N}{2}$$

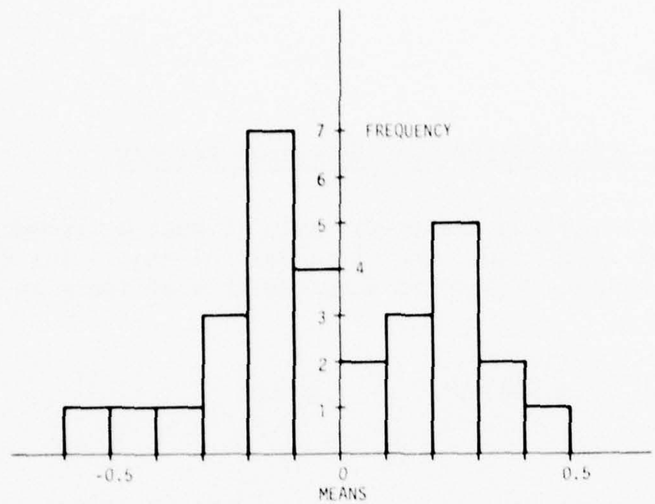
$$\text{standard deviation} \quad \sim \sqrt{\bar{\Delta}^2 \,/\, \frac{N}{2}}$$

and these are the values that we are investigating i.e. the parameters of the distribution of the difference of the sample means.

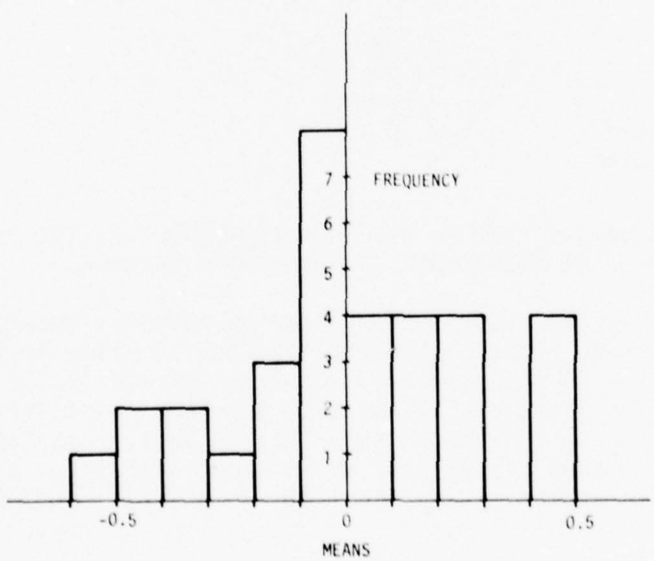
Consider the rectangular distribution, mean ½, variance about the mean 1/12. From Table 1 we should expect to obtain the results given in Table 4 and this is consistent with the numerical results obtained (Table 2). Numerical results obtained for the means relative to the gaussian distribution also compare favourably with analytical results. In fact, from Table 3, the means for the random and bialternate methods are near zero, as they should be. From Appendix D, an approximation for $|\bar{\Delta}|$ in the alternate method is

$$N \sqrt{\frac{2}{\pi}} \times 0.928 \times \left(\frac{N}{2} + \frac{\pi}{8}\right)^{-\pi/4}$$

Evaluating for $N = 4$, $N = 8$ and $N = 20$ we obtain 1.49, 1.84, 2.34 respectively, which give means of 0.75, 0.46 and 0.23.

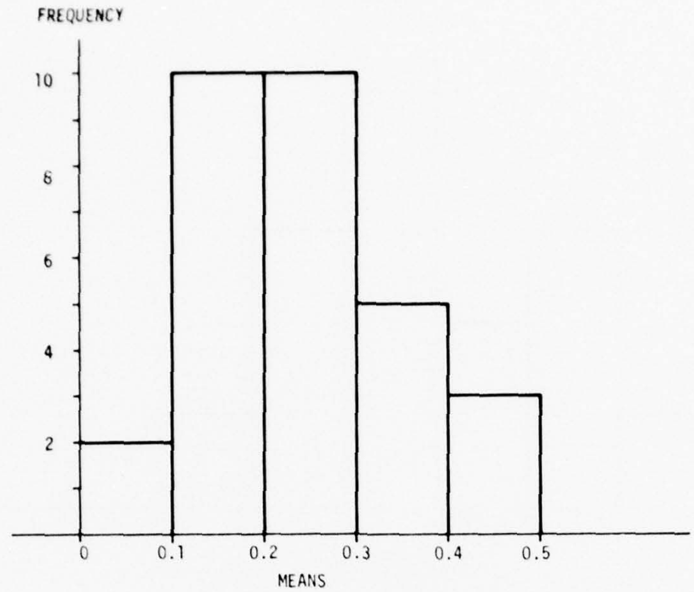


a) N = 4

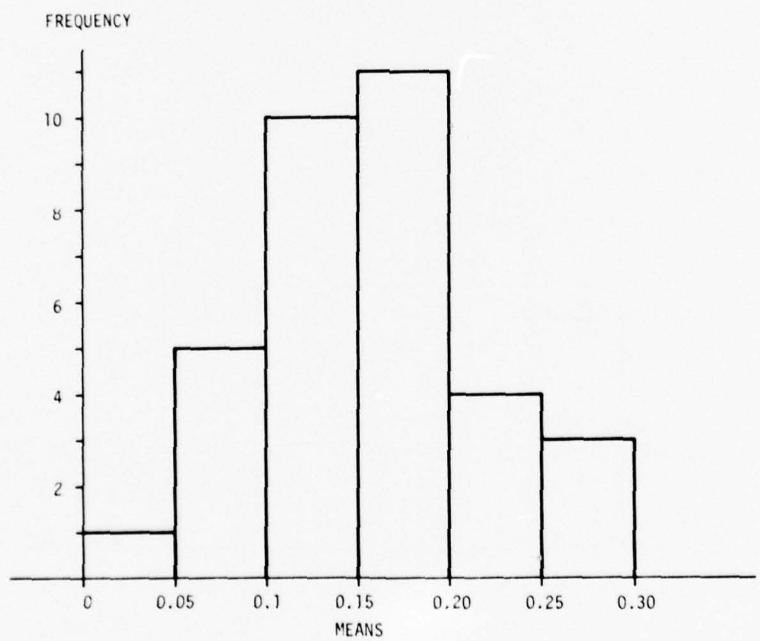


b) N = 6

FIG. 3 UNIFORM DISTRIBUTION
RANDOM METHOD

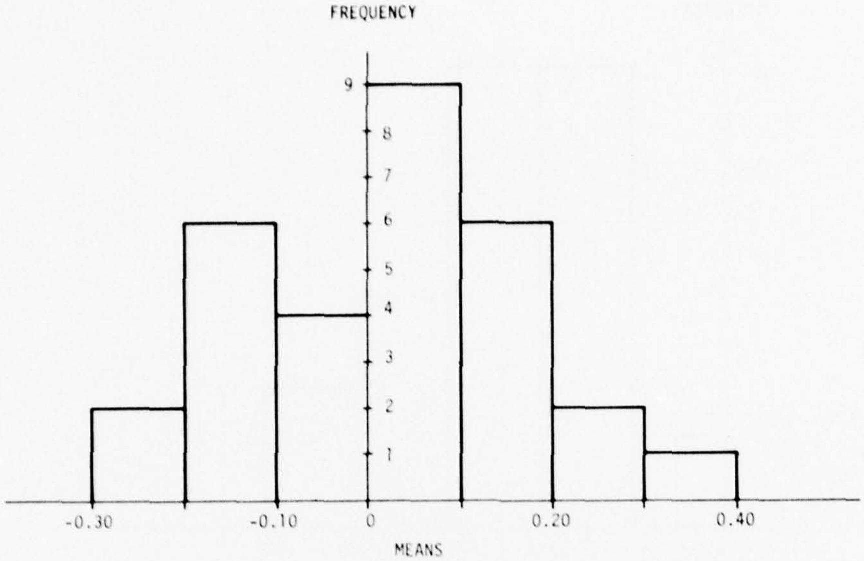


a) N = 4

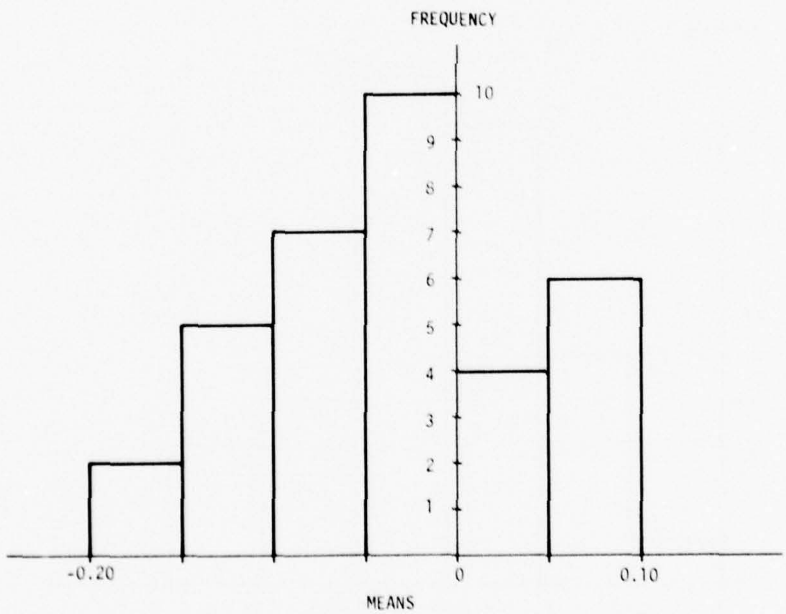


b) N = 6

FIG. 4 UNIFORM DISTRIBUTION
ALTERNATE METHOD



a) N = 4



b) N = 6

FIG. 5 UNIFORM DISTRIBUTION
BIALTERNATE METHOD

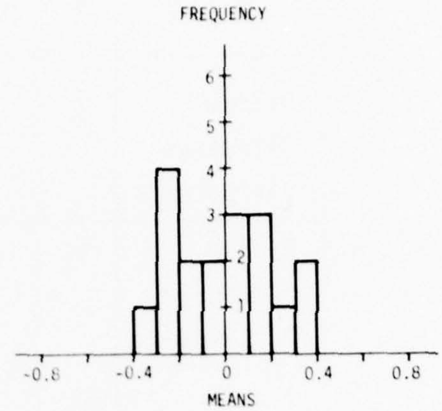


FIG. 6
GAUSSIAN DISTRIBUTION
RANDOM METHOD

N = 4

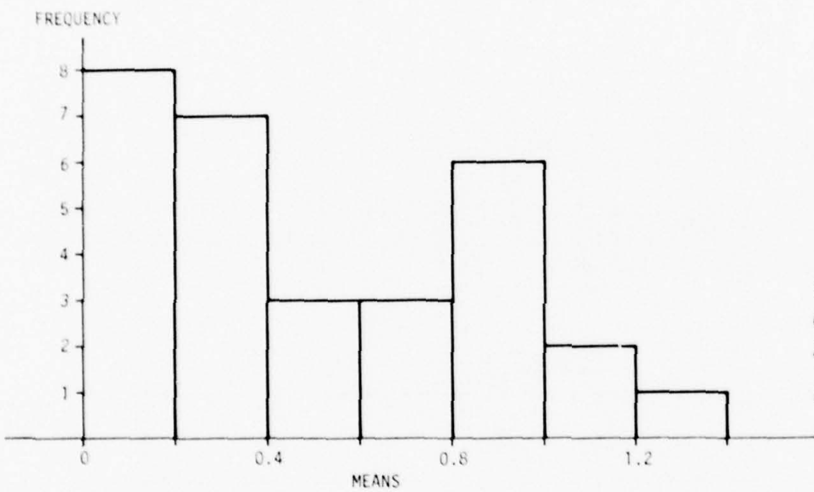


FIG. 7
GAUSSIAN DISTRIBUTION
ALTERNATE METHOD

N = 4

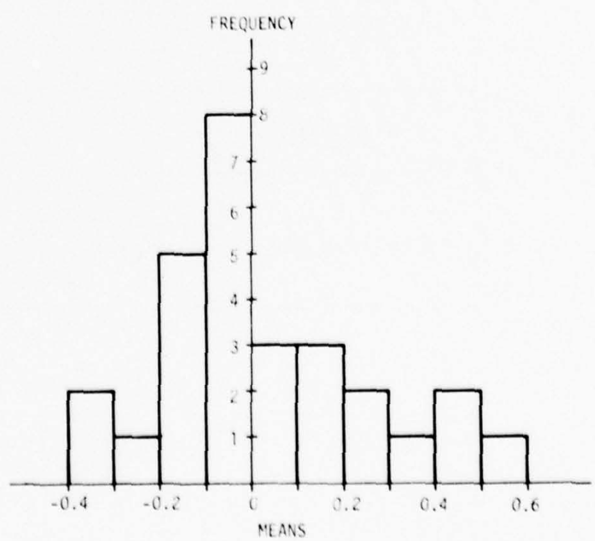


FIG. 8
GAUSSIAN DISTRIBUTION
BIALTERNATE METHOD

N = 4

TABLE 4

METHOD	N	MEAN = $\bar{\Delta}/ N/2$	STANDARD DEVIATION = $\frac{\sqrt{\overline{\Delta^2}}}{N/2}$
Random	4	0	0.29
Alternate		-0.20	0.22
Bialternate		0	0.16
Random	6	0	0.24
Alternate		-0.14	0.15
Bialternate		-0.05	0.15
Random	20	0	0.13
Alternate		-0.05	0.05
Bialternate		0	0.02

2.2 Generalizing the Rectangular Distribution

Ta[illegible]dicated results for a parent distribution with a mean of ½ and standa[illegible]d deviation of $1/\sqrt{12}$. These can be generalized to a standard deviation of σ, as shown in Table 5.

TABLE 5

METHOD	$\bar{\Delta}$	$\overline{\Delta^2}$	$\sigma_{\bar{\Delta}}$	
Random	0	$N\sigma^2$	$\sigma\sqrt{N}$	
Alternate	$\frac{-N\sqrt{3}\,\sigma}{N+1}$	$3N\sigma^2/(N+1)$	$\sqrt{3N}\,\sigma/N+1$	
Bialternate	0	$6\sigma^2/(N+1)$	$\sigma\sqrt{\frac{6}{N+1}}$	$\frac{N}{2}$ even
Bialternate	$-\frac{2\sqrt{3}\,\sigma}{N+1}$	$18\sigma^2/(N+1)$	$\sigma\sqrt{\frac{6(3N+1)}{N+1}}$	$\frac{N}{2}$ odd

2.3 The Frequency Distributions of $\Delta/N/2$

Figures 3 to 8 give an idea of the frequency distributions of $\Delta/N/2$ i.e. the frequency distribution of the difference of the sample means for various values of N ($\bar{\Delta}$ in the alternate ranking method is taken to be positive) and they will be of the types shown in Fig. 9.

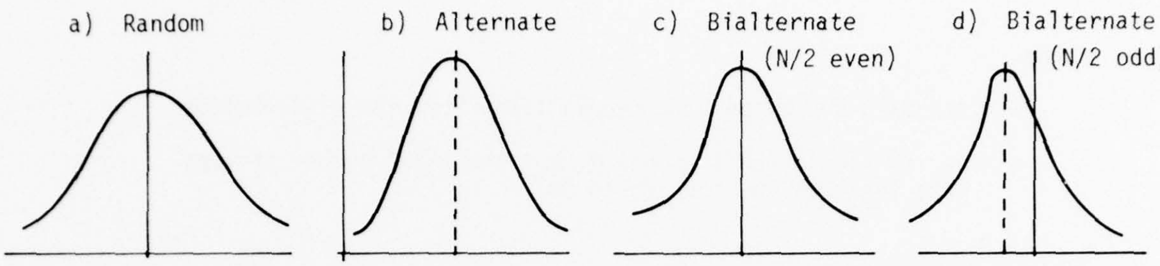


FIG. 9 PARENT POPULATION RECTANGULAR

Alternate b) and Bialternate d) show a displaced mean. From Fig. 1 both these tend to 0 for $N \to \infty$, in b) as $1/N$ and in d) as $1/N^2$.

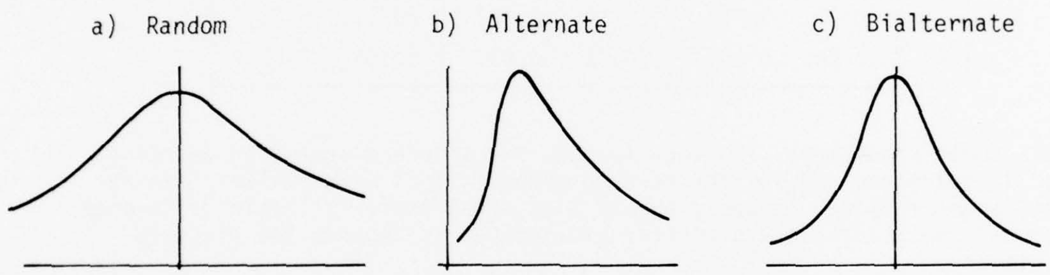


FIG. 10 PARENT POPULATION GAUSSIAN

These shows the advantages of the bialternate ranking method. It gives a distribution closely centred about a mean that is either 0 or almost zero. Taking into consideration the numerical results for $N = 20$, if the parent population is gaussian the bialternate ranking method gives a standard deviation of 0.05. Samples of 1600 items would have to be taken in order to have such a result using the random method.

Figure 2a shows that when the parent population is rectangular $\bar{\Delta}^2$ increases linearly. This means a standard deviation that decreases as $1/\sqrt{N}$. Figure 2b shows that in the alternate ranking process the standard deviation $\to 0$ as $N \to \infty$ as $1/N$ while Figs. 2c & d show that in the bialternate method the standard deviation $\to 0$ as $1/N\sqrt{N}$, which are much more desirable results.

Not having theoretical results for the standard deviation in the case of a Gaussian parent population, it is difficult to say what will happen when N is large.

An indication of the real value of the standard deviation may be had by calculating confidence intervals. This can be done using χ^2. For confidence intervals with a level of confidence of 95%

$$\frac{S\sqrt{n}}{\chi_{0.975}} < \sigma < \frac{S\sqrt{n}}{\chi_{0.025}} ,$$

where

s = standard deviation of a sample taken from the distribution

n = no. in sample (this is not N, but the total number of runs done for each N i.e. 20 to 30).

Table 6 is obtained:

TABLE 6

N	RANDOM	ALTERNATE	BIALTERNATE
4	(0.49, 0.84)	(0.28, 0.48)	(0.27, 0.46)
8	(0.38, 0.74)	(0.09, 0.18)	(0.11, 0.21)
20	(0.26, 0.50)	(0.04, 0.08)	(0.04, 0.08)

This table shows that, for each method, the standard deviation decreases as N increases and for the ranking methods it is much smaller than for the random method. It would appear also, that there is little difference between the alternate and bialternate methods as regards the standard deviation.

REFERENCE

1. ABRAMOWITZ, M. and STEGUN, I.A. eds. Handbook of Mathematical Functions, with formulas, graphs and mathematical tables. Wash., National Bureau of Standards, 1964. Reprinted with corrections 1968.

APPENDICES

APPENDIX A

CALCULATION OF $\overline{\Delta}$ FOR A POPULATION OF PROBABILITY DISTRIBUTION $p(x)$

Consider a population of probability distribution $p(x)$. Take N samples of values $x_1\ x_2 \ldots x_N$. The probability that one of these lies between x_r and $x_r + dx_r$ is $p(x_r)\,dx_r$ and therefore the chance that the first item is at x_1, the second at x_2 etc... is $p(x_1)\,p(x_2) \ldots p(x_N)\,dx_1 \ldots dx_N$. But there are $N!$ permutations of the Nx's so that the probability of a random sample of N having these values is

$$N!\ p(x_1)\,p(x_2) \ldots p(x_N)\,dx_1 \ldots dx_N$$

Ordering the random sample $x_1 \le x_2 \le \ldots \le x_N$ .. 1. To calculate $\overline{\Delta}$, $\sum_{r=1}^{N} x_r \gamma_r$ must be integrated over all values of x_r subject to condition 1.

$$\overline{\Delta} = N! \int\int \ldots \int p(x_1) \ldots p(x_N) \sum_{r=1}^{N} x_r \gamma_r\, dx_1 \ldots dx_N .$$

Choose the order of integration for the term $\gamma_r x_r$ as $(dx_1\, dx_2 \ldots dx_{r-1})\ (dx_{r+1}\, dx_{r+2} \ldots dx_N)\ dx_r$ and so

$$\overline{\Delta} = N! \sum_{r=1}^{N} \int_{-\infty}^{+\infty} I_r J_r \gamma_r x_r\, p(x_r)\, dx_r$$

where

$$I_r = \int_{-\infty}^{x_r} p(x_r-1)\, dx_{r-1} \int_{-\infty}^{x_{r-1}} p(x_{r-2})\, dx_{r-2} \ldots \int_{-\infty}^{x_2} p(x_1)\, dx_1$$

and

$$J_r = \int_{x_r}^{\infty} p(x_{r+1})\, dx_{r+1} \int_{x_{r+1}}^{\infty} p(x_{r+2})\, dx_{r+2} \ldots \int_{x_{N-1}}^{\infty} p(x_N)\, dx_N$$

Lemma 1

$$I_r = \frac{P(x_r)^{r-1}}{(r-1)!}$$

Proof

$$I_2 = \int_{-\infty}^{x_2} p(x_1)\, dx_1 = P(x_2) \implies \text{result true for } r = 2$$

generally assume

$$I_s = \frac{P(x_s)^{s-1}}{(s-1)!}$$

Then

$$I_{s+1} = \int_{-\infty}^{x_{s+1}} \frac{p(x_s)\, P(x_s)^{s-1}}{(s-1)!}\, dx_s$$

$$= \left[\frac{P(x_s)\, P^{s-1}(x_s)}{(s-1)!}\right]_{-\infty}^{x_{s+1}} - \frac{(s-1)}{(s-1)!} \int_{-\infty}^{x_{s+1}} P(x_s)\, P(x_s)^{s-2}\, p(x_s)\, dx_s$$

$$= \frac{P(x_{s+1})^s}{(s-1)!} - \frac{(s-1)}{(s-1)!} \int_{-\infty}^{x_{s+1}} p(x_s)\, P^{s-1}(x_s)\, dx_s$$

$$= \frac{P(x_{s+1})^s}{(s-1)!} - (s-1)\, I_{s+1}$$

$$\implies s\, I_{s+1} = \frac{P(x_{s+1})^s}{(s-1)!}$$

$$\implies I_{s+1} = \frac{P(x_{s+1})^s}{s!}$$

so by induction true for all s .

Lemma 2

$$J_r = \frac{Q^{N-r}(x_r)}{(N-r)!}$$

Proof

$$J_{N-1} = \int_{x_{N-1}}^{\infty} p(x_N)\, dx_N = Q(x_{N-1}) \implies \text{lemma true for } r = N-1$$

Assume

$$J_s = \frac{Q^{N-s}(x_s)}{(N-s)!}$$

$$J_{s-1} = \int_{x_{s-1}}^{\infty} P(x_s) \frac{Q^{N-s}(x_s)}{(N-s)!}\, dx_s$$

$$= \frac{Q(x_{s-1})^{N-s+1}}{(N-s)!} - \int_{x_{s-1}}^{\infty} \frac{Q(x_s)}{(N-s)!}(N-s)\, Q^{N-s-1}(x_s)\, P(x_s)\, dx_s$$

$$= \frac{Q(x_{s-1})^{N-s+1}}{(N-s)!} - \int_{x_{s-1}}^{\infty} (N-s)\, P(x_s) \frac{Q^{N-s}(x_s)}{(N-s)!}\, dx_s$$

$$= \frac{Q(x_{s-1})^{N-s+1}}{(N-s)!} - (N-s)\, J_{s-1}$$

$$(N-s+1)\, J_s = \frac{Q(x_{s-1})^{N-s+1}}{(N-s)!}$$

$$\implies J_s = \frac{Q(x_{s-1})^{N-s+1}}{(N-s+1)!}$$

Applying the lemmas

$$\bar{\Delta} = N! \sum_{r=1}^{N} \int_{-\infty}^{+\infty} \gamma_r x_r P(x_r) \frac{P^{r-1}(x_r)}{(r-1)!} \frac{Q^{N-r}(x_r)}{(N-r)!}\, dx_r$$

Replace x_r by y and $P = P(y)$, $p = p(y)$

so

$$\bar{\Delta} = N! \sum_{r=1}^{N} \int_{-\infty}^{+\infty} \gamma_r\, y\, p \frac{P^{r-1}}{(r-1)!} \frac{Q^{N-r}}{(N-r)!}\, dy$$

APPENDIX B

CALCULATION OF $\bar{\Delta}^2$ FOR A POPULATION OF PROBABILITY DISTRIBUTION $p(x)$

$$\bar{\Delta}^2 = N! \int\int \ldots \int \left[\sum_{r=1}^{N} \gamma_r x_r \right]^2 p(x_1) \ldots p(x_N)\, dx_1 \ldots dx_N$$

since $\gamma_r^2 = 1$, expanding the squared term

$$\bar{\Delta}^2 = N! \int\int \ldots \int \left[\sum_{r=1}^{N} x_r^2 + 2 \sum_{r=1}^{N-1} \sum_{s>r}^{N} \gamma_r \gamma_s x_r x_s \right] p(x_1) \ldots p(x_N)\, dx_1 \ldots dx_N$$

Consider these terms separately. For the integral of Σx_r^2 , the same technique as for $\bar{\Delta}$ may be used leading to

$$N! \int_{-\infty}^{+\infty} \sum_{r=1}^{N} p y^2 \frac{P^{r-1} \quad Q^{N-r}}{(r-1)! \, (N-r)!} dy$$

But since the γ_r terms do not appear, the summation becomes a single binomial

$$(P+Q)^{N-1} = \sum_{r=1}^{N} \frac{(N-1)!}{(r-1)!\,(N-r)!} P^{r-1} Q^{N-r} \quad ,$$

whence this term becomes

$$N \int_{-\infty}^{+\infty} p y^2 \, dy = N v^2 \quad ,$$

where v^2 is the variance about the origin of the original population.

For the second integral, integrate in the order

$$(dx_1\, dx_2 \ldots dx_{r-1})\,(dx_{r+1}\, dx_{r+2} \ldots dx_{s-1})\,(dx_{s+1}\, dx_{s+2} \ldots dx_N)\, dx_r\, dx_s \ .$$

So the integral becomes

$$2 \sum_{r=1}^{N-1} \sum_{s=r+1}^{N} N! \int_{-\infty}^{+\infty} x_s\, p(x_s)\, dx_s \int_{-\infty}^{x_s} x_r\, p(x_r)\, dx_r \; I_r J_s K_{rs} \quad ,$$

PRECEDING PAGE BLANK

where I_r, J_s have the meanings previously used while

$$K_{rs} = \int_{x_r}^{x_s} p(x_{s-1})\, dx_{s-1} \int_{x_r}^{x_{s-1}} p(x_{s-2})\, dx_{s-2} \cdots \int_{x_r}^{x_{r+2}} p(x_{r+1})\, dx_{r+1}$$

Now

$$K_{r,r+2} = \int_{x_r}^{x_{r+2}} p(x_{r+1})\, dx_{r+1} = P(x_{r+2}) - P(x_r)$$

$$K_{r,r+3} = \int_{x_r}^{x_{r+3}} [P(x_{r+2}) - P(x_r)]\, p(x_{r+2})\, dx_{r+2}$$

$$= \frac{[P(x_{r+2}) - P(x_r)]^2}{2} \Bigg|_{x_r}^{x_{r+3}} = \frac{[P(x_{r+3}) - P(x_r)]^2}{2}$$

and generally

$$K_{rs} = \frac{[P(x_s) - P(x_r)]^{s-r-1}}{(s-r-1)!}$$

Hence the second integral becomes

$$2N! \sum_{r=1}^{N-1} \sum_{s=r+1}^{N} \int_{-\infty}^{+\infty} x_s\, p(x_s)\, dx_s \int_{-\infty}^{x_s} x_r\, p(x_r)\, dx_r \frac{P^{r-1}(x_r)}{(r-1)!} \frac{Q^{N-s}(x_s)}{(N-s)!} \frac{[P(x_s)-P(x_r)]^{s-r-1}}{(s-r-1)!}$$

replace x_s with y and x_r with z so

$$\overline{\Delta}^2 = Nr^2 + 2N! \sum_{r=1}^{N-1} \sum_{s=r+1}^{N} \int_{-\infty}^{+\infty} y\, Py\, dy \int_{-\infty}^{y} z\, Pz\, dz\, \frac{P^{r-1}(z)}{(r-1)!} \frac{Q^{N-s}(y)}{(N-s)!} \frac{[P(y) - P(z)]^{s-r-1}}{(s-r-1)!}$$

APPENDIX C

CALCULATION OF $\overline{\Delta}$ AND $\overline{\Delta}^2$ WHERE PARENT POPULATION IS RECTANGULAR AND THE METHOD USED IS THE BIALTERNATE RANKING METHOD

An expression for γ_r is given by

$$\gamma_r = \sqrt{2} \cos [\pi/4 + (r-1)\,\pi/2]$$
$$= \sqrt{2}\ \mathrm{Re}\ e^{i\pi/4}\, e^{i\pi/2(r-1)}$$

where Re = real part of the complex expression

$$\overline{\Delta} = N!\ \mathrm{Re}\sqrt{2} \sum_{r=1}^{N} \int_{-\infty}^{+\infty} p(y)\ y\, e^{i\pi/4} \frac{\left(P(y)\, e^{i\pi/2}\right)^{r-1}}{(r-1)!} \frac{Q(y)^{N-r}}{(N-r)!}\, dy$$

but

$$p(y) = \begin{cases} 1 & y \in [0,1] \\ 0 & y \notin [0,1] \end{cases}$$

$$P(y) = y\ ; \quad Q(y) = 1 - y$$

$$\Rightarrow \overline{\Delta} = N!\ \mathrm{Re}\sqrt{2} \sum_{r=1}^{N} \int_0^1 y\, e^{i\pi/4} \frac{\left(y\, e^{i\pi/2}\right)^{r-1}}{(r-1)!} \frac{(1-y)^{N-r}}{(N-r)!}\, dy$$

$$= N\ \mathrm{Re}\sqrt{2}\, e^{i\pi/4} \int_0^1 y(1-y+y\, e^{i\pi/2})^{N-1}\, dy$$

$$e^{i\pi/4} = 1/\sqrt{2}\,(1+i)$$

$$\Rightarrow \overline{\Delta} = N\ \mathrm{Re}(1+i) \int_0^1 y(1+\alpha y)^{N-1}\, dy$$

where

$$\alpha = e^{i\pi/2} - 1 = i - 1$$

Integrating by parts

$$\bar{\Delta} = N\ Re(1+i)\left\{\left[\frac{y(1+\alpha y)^N}{\alpha N}\right]_0^1 - \frac{1}{\alpha N}\int_0^1 (1+\alpha y)^N\ dy\right\}$$

$$= N\ Re(1+i)\left[\frac{(1+\alpha)^N}{\alpha N} - \frac{(1+\alpha y)^{N+1}}{\alpha^2 N(N+1)}\Bigg|_0^1\right]$$

Since N is even

$$\bar{\Delta} = N\ Re(1+i)\left[\frac{(-1)^{N/2}}{\alpha N} - \frac{i(-1)^{N/2}}{\alpha^2 N(N+1)} + \frac{1}{\alpha^2 N(N+1)}\right]$$

$$\frac{1}{\alpha} = \frac{1}{i-1} = \frac{i+1}{-2} \qquad\qquad \frac{1}{\alpha^2} = \frac{2i}{4} = \frac{i}{2}$$

$$\Rightarrow \bar{\Delta} = N\ Re(1+i)\left[\frac{(-1)^{N/2+1}\ (i+1)}{2N} + \frac{(-1)^{N/2}}{2N(N+1)} + \frac{i}{2N(N+1)}\right]$$

$$= N\ Re\left[\frac{i(-1)^{N/2+1}}{N} + \frac{(-1)^{N/2}(1+i)}{2N(N+1)} + \frac{(i-1)}{2N(N+1)}\right]$$

$$= \frac{(-1)^{N/2}}{2(N+1)} - \frac{1}{2(N+1)} = \frac{(-1)^{N/2} - 1}{2(N+1)}$$

If $N/2$ is even $\bar{\Delta} = 0$

If $N/2$ is odd $\bar{\Delta} = -\dfrac{1}{N+1}$

To find $\bar{\Delta}^2$ an expression for $\gamma_r\ \gamma_s$ must be found

$$\gamma_r\ \gamma_s = 2\cos(\pi/4 + (r-1)\ \pi/2)\cos(\pi/4 + (s-1)\pi/2)$$

$$= \cos(\pi/2 + (r+s-2)\ \pi/2) + \cos(s-r)\ \pi/2$$

$$= \cos(s+r-1)\ \pi/2 + \cos(s-r)\ \pi/2$$

$$= Re\left[e^{i\pi/2(s+r-1)} + e^{i\pi/2(s-r)}\right]$$

so

$$\bar{\Delta}^2 = N/3 + N!\ \mathrm{Re} \sum_{r=1}^{N-1} \sum_{s=r+1}^{N} \int_0^1 y\,dy \int_0^y z\,dz \left(e^{i\pi/2(s+r-1)} + e^{i\pi/2(s-r)}\right)$$

$$\cdot \frac{z^{r-1}(1-y)^{N-s}(y-z)^{s-r-1}}{(r-1)!\ (N-s)!(s-r-1)!}$$

Now

$$e^{i\pi/2(s+r-1)} = e^{i\pi(r-1)}\ e^{i\pi/2(s-r-1)}\ e^{i\pi}$$

and

$$e^{i\pi/2(s-r)} = e^{i\pi/2(s-r-1)}\ e^{i\pi/2}$$

$$\Rightarrow \bar{\Delta}^2 = N/3 + N!\ \mathrm{Re} \sum_{r=1}^{N-1} \sum_{s=r+1}^{N} \int_0^1 y\,dy \int_0^y z\,dz \quad \cdot$$

$$\left\{ \frac{e^{i\pi}(z\,e^{i\pi})^{r-1}\ (1-y)^{N-s}\left((y-z)\ e^{i\pi/2}\right)^{s-r-1}}{(r-1)!\ (N-s)!\ (s-r-1)!} + \right.$$

$$\left. \frac{e^{i\pi/2}\ z^{r-1}(1-y)^{N-s}\ \left((y-z)\ e^{i\pi/2}\right)^{s-r-1}}{(r-1)!\ (N-s)!\ (s-r-1)!} \right\}$$

which using the multinominal theorem becomes

$$\bar{\Delta}^2 = N/3 + 2N(N-1)\ \mathrm{Re} \int_0^y y\,dy \left[e^{i\pi}\left(1+y(e^{i\pi/2}-1)+z(e^{i\pi}-e^{i\pi/2})\right)^{N-2} + \right.$$

$$\left. e^{i\pi/2}\left(1+y(e^{i\pi/2}-1) + z(1-e^{i\pi/2})\right)^{N-2}\right] z\,dz$$

$$e^{i\pi} = -1 \qquad e^{i\pi/2} = i$$

$$\Rightarrow \bar{\Delta}^2 = N/3 + 2N(N-1)\,\mathrm{Re} \int_0^1 y\,dy \int_0^y z\,dz \Big[-(1+y(i-1)-z(i+1))^{N-2} + i[1+y(i-1)+z(1-i)]^{N-2}\Big]$$

The first term is

$$-2N(N-1)\,\mathrm{Re} \int_0^1 y\,dy \int_0^y z\,dz\,[1+y(i-1)-z(i+1)]^{N-2}$$

Integrating over z by parts gives

$$-2N(N-1)z\frac{[1+y(o-1)-z(i+1)]^{N-1}}{-(i+1)N}\Bigg|_o^y \frac{-2(N-1)}{(1+i)}\int_0^y \Big(1+y(i-1)-z(i+1)\Big)^{N-1} dz$$

$$= (N-1)(1-i)(1-2y)^{N-1} - i(N-1)\,[(1-2y)^N - (1+y(i-1))^N]$$

The integration over y for the first term gives

$$(N-1)(1-i)\int_0^1 y(1-2y)^{N-1}\,dy = (-1)^{N+1}(1-i)\,\frac{(N-1)}{2N} - \frac{(N-1)(1-i)}{4N(N+1)}[(-1)^{N+1}-1]$$

N is even and only the real part is required. This is $\dfrac{-(N-1)}{2(N+1)}$

The second term is

$$2iN(N-1)\int_0^1 y\,dy \int_0^y z\,dz\,[1+y(i-1)+z(1-i)]^{N-2}$$

Integration over z gives

$$Ny(i-1)+1 - [1+y(i-1)]^N$$

$\Rightarrow$ Integrating this over y and taking the real part gives

$$-\frac{N}{3} + \frac{1}{2} + \frac{(-1)^{N/2+1}}{2(N+1)}$$

$$\bar{\Delta}^2 = N/3 - \frac{(N-1)}{2(N+1)} - \frac{N}{3} + \frac{1}{2} + \frac{(-1)^{N/2+1}}{2(N+1)} = \frac{2+(-1)^{N/2+1}}{2(N+1)}$$

If $N/2$ is even $\bar{\Delta}^2 = \frac{1}{2}(N+1)$

If $N/2$ is odd $\bar{\Delta}^2 = \frac{3}{2}(N+1)$

APPENDIX D

AN APPROXIMATION OF $\bar{\Delta}$ WHERE THE ALTERNATE RANKING METHOD IS USED AND THE PARENT POPULATION IS GAUSSIAN

The integral to approximate is

$$\bar{\Delta} = 2N \int_0^\infty y \left(p(y)\right) (P-Q)^{N-1} dy$$

$$= 2N \int_0^\infty y \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} (2P-1)^{N-1} dy$$

This was done both numerically and theoretically. The numerical approximation was obtained using the generalized Simpson Rule and results were

$N = 2$ $\quad \bar{\Delta} = 1.13$

$N = 4$ $\quad \bar{\Delta} = 1.47$

For the theoretical approximation use p. 933 of Ref. 1

$$P(x) \sim \tfrac{1}{2} + \tfrac{1}{2}\left(1 - e^{-\frac{2y^2}{\pi}}\right)^{\frac{1}{2}}$$

So

$$\bar{\Delta} = N\sqrt{\frac{2}{\pi}} \int_0^{+\infty} y\, e^{-y^2/2}\left(1-e^{-\frac{2y^2}{\pi}}\right)^{N-1/2} dy$$

change the variable $\quad e^{-\frac{2y^2}{\pi}} = t$

so $\quad y\,dy = \dfrac{-\pi}{4}\,\dfrac{dt}{t}$

Hence

$$\bar{\Delta} = N\sqrt{\frac{2}{\pi}}\,\frac{\pi}{4} \int_0^1 t^{(\pi/4-1)}(1-t)^{N-1/2}\, dt$$

and this integral is equal to the beta function

$$B\left(\frac{\pi}{4}, \frac{N+1}{2}\right) = \Gamma\left(\frac{\pi}{4}\right)\Gamma\left(\frac{N+1}{2}\right) \qquad \text{(p. 258 of Ref. 1).}$$

At this point the tabulated Γ values can be used or else a further approximation using Sterling's formule (p. 257 of Ref. 1)

$$\lg\left(\Gamma\left(\frac{N+1}{2}\right)\right) \sim N/2 \lg\left(\frac{N+1}{2}\right) - \frac{N+1}{2} + \frac{1}{2}\lg 2\pi + \frac{1}{12\,\frac{(N+1)}{2}} + 0\,\frac{1}{(N+1)^3}$$

$$\lg\left(\Gamma\left(\frac{N}{2} + \frac{1}{2} + \frac{\pi}{4}\right)\right) \sim \left(\frac{N}{2} + \frac{\pi}{4}\right)\lg\left(\frac{N+1+\pi/2}{2}\right) - \frac{N+1}{2} - \frac{\pi}{4} + \frac{1}{2}\lg 2\pi$$

$$+ \frac{1}{12\,\frac{(N+1+\pi/2)}{2}} + \frac{1}{0(N^3)}$$

Hence

$$\lg\left[\frac{\Gamma\left(\frac{N+1}{2}\right)}{\Gamma\,\frac{N+1}{2} + \frac{\pi}{4}}\right] \sim \frac{N}{2}\lg\left(\frac{N+1}{2}\right) - \left(\frac{N}{2} + \frac{\pi}{4}\right)\lg\left(\frac{N+1+\pi/2}{2}\right)$$

$$+ \frac{\pi}{4} + \frac{\pi/2}{6(N+1)(N+1+\pi/4} + 0\left(\frac{1}{N^3}\right)$$

Now write $N = u + \gamma$ where γ is a constant to be determined. The reason for this is that, as will be seen, by suitably choosing γ, the logarithms may be expanded to give the best available accuracy. Retaining only terms of the order of $1/N$

$$\lg\left[\frac{\Gamma\left(\frac{N+1}{2}\right)}{\Gamma\,\frac{N+1}{2} + \frac{\pi}{4}}\right] \sim \frac{u+\gamma}{2}\lg\frac{u+\gamma+1}{2} - \frac{u+\gamma+\pi/2}{2}\lg\frac{u+\gamma+1+\pi/2}{2} + \frac{\pi}{4}$$

$$= \frac{u+\gamma}{2}\left[\lg\frac{u}{2} + \lg\left(1+\frac{\gamma+1}{u}\right)\right] - \frac{u+\gamma+\pi/2}{2}\left[\lg\frac{u}{2} + \lg\left(1+\frac{\gamma+1+\pi/2}{4}\right)\right] + \frac{\pi}{4}$$

Expanding the lg (1+...) terms for large enough u

$$= -\frac{\pi}{4}\lg\frac{u}{2} + \left(\frac{u+\gamma}{2}\right)\left[\frac{\gamma+1}{u} - \frac{(\gamma+1)^2}{2u^2} + 0\left(\frac{1}{u^3}\right)\right]$$

$$- \frac{(u+\gamma+\pi/2)}{2}\left[\frac{\gamma+1+\pi/2}{u} - \frac{(\gamma+1+\pi/2)^2}{2u^2} + 0\left(\frac{1}{u^3}\right)\right] + \frac{\pi}{4}$$

$$= -\frac{\pi}{4}\lg u/2 + 1/4u\,(-\pi\gamma - \pi^2/4)$$

So the term in $1/u$ is identically zero if $\gamma = -\pi/4$ so that $N = u-\pi/4$ and the lg term becomes $-\pi/4 \lg (N/2 + \pi/8)$ i.e.

$$\frac{\Gamma\left(\frac{N+1}{2}\right)}{\Gamma\left(\frac{N+1}{2}+\frac{\pi}{4}\right)} \sim \left(\frac{N}{2}+\frac{\pi}{8}\right)^{-\pi/4}$$

so

$$\bar{\Delta} \sim N/2 \sqrt{\pi/2}\ \Gamma(\pi/4)\ (N/2 + \pi/8)^{-\pi/4}$$

also $\quad \Gamma(1+\pi/4) = \pi/4\ \Gamma(\pi/4)$

and

$$\Gamma(1+\pi/4) = 0.928$$

$$\Rightarrow \bar{\Delta} = N/2 \sqrt{\pi/2} \times 4/\pi \times 0.928 \times (N/2 + \pi/8)^{-\pi/4}$$

$$= N \sqrt{2/\pi} \times 0.928 \times (N/2 + \pi/8)^{-\pi/4}$$

APPENDIX E

A COMMENT ON LINEAR PROGRAMMING

It has already been said that linear programming cannot be used when only ranking data are available; but even if there are sufficient data the technique may not produce satisfactory results. First of all, the solution may not be unique and so a criterion has to be found for selecting the best of the optimal solutions. Secondly, the linear program minimizes the difference of the means of the two samples, but this may be at the cost of an even spread of the items in the two samples i.e.

With a linear program the following optimal solution is found

A	B
0.01	0.30
0.01	0.37
0.02	0.40
0.04	0.44
0.84	0.51
0.89	0.56
0.93	0.57
0.99	0.58

(samples taken from a rectangular distribution mean ½ , variance about mean 1/12)

The difference of the means is 0 but sample A contains small values and large ones while B contains all intermediate values. The ranking methods assured a more even spread of values in the samples although the mean may not be 0 , but generally speaking this situation is more desirable.

PRECEDING PAGE BLANK